人机合作使人更冒险：主体责任感的中介作用*

耿晓伟[1，2] 刘超[3] 苏黎[1] 韩冰雪[1] 张巧明[4] 吴明证[5]

（[1]杭州师范大学心理系，杭州，311121）

（[2]浙江省哲学社会科学培育实验室“杭州师范大学婴幼儿发展与托育实验室”，杭州 311121）

（[3]滨州职业学院马克思主义学院，滨州 256600）（[4]鲁东大学教育学院，烟台 264025）

（[5]浙江大学心理与行为科学系，杭州 310028）

摘要　随着人工智能技术的迅猛发展，人工智能越来越成为人的“助攻”。在人机合作风险决策的过程中，人工智能是否会助长人类的冒险行为，以及人知觉到的主体责任如何发挥作用，这些问题亟待澄清。为了考察人-机合作对个体风险决策的影响及其机制，进行了四个实验。结果发现：（1）不管与人合作，还是与人工智能合作，个体都比单独做决策时更保守；“人-机”合作比“人-人”合作时，个体更冒险。（2）个体在合作中知觉到的主体责任部分中介了“人-机”合作对风险决策的影响，人机合作时，个体知觉到的主体责任更大，从而在风险决策中更冒险。（3）成功反馈时，人机合作情景下，个体更多将责任归于自己，知觉到的主体责任在人机合作对风险决策影响中起中介作用；失败反馈时，人机合作与“人-人”合作之间知觉到的主体责任差异不显著，知觉到的主体责任的中介作用不成立。

关键词:“人-机”合作；“人-人”合作；风险决策；知觉到的主体责任；结果反馈

1 引言

随着人工智能时代的到来，人工智能广泛应用于涉及风险决策的领域，如自动驾驶、医疗决策、投资决策等。在许多场景中，人类将与人工智能联合行动、共同决策。人工智能充当的角色不再仅是工具，而成为人类的“助攻”。例如，在多人在线战斗竞技游戏比赛中，人工智能“助攻”能够快速分析游戏地图中的各种信息，然后根据这些信息做出决策，如提醒人类何时躲避敌方的偷袭。同时，人类玩家也会向人工智能 “助攻” 传达自己的意图，比如，想要进攻某一路线。人工智能“助攻”就能结合自己分析的数据，提出更合理的进攻策略，人类玩

* 收稿日期：2024-02-08
国家自然科学基金项目（71971104）资助。
通信作者：耿晓伟，E-mail：xwgeng@hznu.edu.cn；吴明证，E-mail：psywu@zju.edu.cn

家和人工智能“助攻”相互配合，共同提高比赛胜率。这种人工智能与人密切合作，以完成联合任务的过程，即人机合作(Lei & Rau, 2021)。在这种人机合作中，人工智能是否会助长人类的冒险行为呢？进一步地，人工智能可能会使个体的自我增强，在人机合作中知觉到的主体责任更强。那么，人工智能是否会通过提升个体知觉到的主体责任，进而助长其冒险行为呢？目前这些问题都还没有答案，亟待澄清。研究人工智能对个体冒险行为的影响及其中介机制有助于更全面地了解人机合作中人工智能对人类行为的影响，包括可能的消极影响，同时为更有效地人机合作决策提供相应的建议。因此，本研究旨在比较“人-机”合作和“人-人”合作对个体风险决策的影响，并探究个人知觉到的主体责任如何发挥作用。

1. 1 人机合作与风险决策

合作通常指两个或两个以上的人或群体为达到共同目的而协同活动，为实现某种特定的目标，心理和行为相互配合、彼此协调共济的一种互动的行为方式 (Decety et al., 2004) 。人机合作指的是人工智能与人密切合作，以完成联合任务 (Lei & Rau, 2021)。人机合作有多种方式，一种合作方式是，人在人机合作中占主导地位，这是最常见的人机合作方式（Duan et al., 2019; Haesevoets et al., 2021）。如人工智能辅助决策，由人来做最后的决定。然而，这种由人主导的人机合作团队往往会受到人的决策偏见（例如过度自信、赌徒谬误等）的影响，难以实现最优决策（Mittal, 2022；Xiong et al., 2023）。与此相反的一种合作方式是，人工智能在人机合作中占主导地位（Yue & Li, 2023）。另一种合作方式是，人和人工智能在合作中具有平等的地位，通过对双方的选择进行平均，来整合双方的决策（Xiong et al., 2023）。随着人工智能不断升级，人与人工智能平等合作方式会越来越受到重视。因此，本研究中人机合作是指人与人工智能具有平等地位的合作。

以往关于“为他人决策”的研究表明，个体为他人决策时比单独自己决策时更加风险回避 (Bolton et al., 2015；Charness & Jackson, 2009；Wang et al., 2018)。Bolton 等人(2015)的研究中，一种条件是被试单独决策，另一种条件是被试为自己和另外一名伙伴做决策，结果发现，在后一种条件下，个体更保守。Charness 等人(2009)研究结果表明当个体代表其他人做出决策时比独自决策时更保守。Wang 等人(2018)比较了为群体决策和为个体决策的差异，结果发现，个体代表群体决策比单独决策更倾向于风险规避。当个体为群体做出决策时，意识到自己对团体收益负有责任使得个体的选择更加谨慎，更加厌恶风险(Atanasov & Kunreuther, 2016)。

以往关于“两人一起决策”的研究表明，个体与他人一起决策时比单独决策时同样会更加风险回避。唐伟超（2017）采用仿真气球任务结果发现，两人共同决策（两个人经过充分的讨论并达成一致意见之后，做出打气或停止打气的选择）、两人轮流决策（两人在无沟通的情况下，轮流做出打气或停止打气的选择）这两种条件下均比独立决策条件下，个人更规避风险。He 等人 (2012)以夫妻为被试，结果发现，在夫妻两人共同决策（对每个选择进行讨论并达成

一致）和单独决策的比较研究中发现，双人共同决策更倾向于风险规避。当个体的决策不仅影响到自己的利益，还会影响到他人的利益时，个体会变的更保守（Bolton et al., 2015）。与单独做决策相比，当与其他人一起合作决策时，个体会因为考虑到对群体/他人的利益的责任，从而更加风险回避。据此，不管是“人-机”合作还是“人-人”合作，个体都会比单独决策时更倾向于风险回避。

假设 1a：个体在单独决策时比“人-人”合作、“人-机”合作时更倾向冒险。

自我归类理论(Self-categorization theory, SCT) (Turner et al., 1987)将自我类别分为三个层次：最高层关注人类与其他种群的区别，中间层关注自我所在的内群体与外群体区别，最低层关注个体作为群体内成员与群体内其他成员的区别。机器人毕竟是与人类不同的实体。因此，人们更多将人类成员知觉为内群体，而将人工智能看成外群体。大量研究表明个体对内群体成员存在偏好，相比于外群体，人们更容易帮助内群体成员(Hewstone et al., 2002; Levine et al., 2005)，观看内群体成员受灾的图片比观看外群体成员受灾的图片更容易有助人意愿(Forgiarini et al., 2011)。一项通过社会困境来探究合作行为的研究发现，尽管人们认为卡通狗计算机代理更可爱和更信任，但是更愿意与真人和类人计算机代理伙伴合作(Parise et al., 1996)。个体更重视人类合作伙伴而不是机器成员（Gombolay et al., 2015）。有研究发现将机器人作为群体成员，会导致内群体认同的下降，而且群体中机器人的数量越多，内群体认同更低 (Savela et al., 2021)。因此，与“人-人”合作相比，当人与人工智能进行合作时，个体对内群体的认同相对要低，个体会较少地考虑对群体和群体成员的责任，故可能会更冒险。

假设 1b: “人-机”合作比“人-人”合作条件下，个体更倾向冒险。

1.2 人机合作中知觉到的主体责任对风险决策的影响

社会认知中主体能动性—社会关系性双重视角模型（Dual Perspective Model of Agency and Communion）认为，对自我、他人和群体的认知可以分为两大类：与主体能动性(agency)有关的（例如目标实现和完成任务功能）、与社会关系性(communion)相关的（例如关系维护和乐于助人、仁慈等社会功能）（Abele & Wojciszke，2014）。据此，本文认为，在人机合作中，个体的责任知觉可以区分为知觉到的主体责任（即在合作中个体对于目标实现和完成任务所承担的责任）和社会关系责任（即在合作中个体为他人所承担的责任）。

群体共同完成任务时，每个群体成员所分担的责任会由于分散到各个成员身上而减小，即责任分散（Darley & Latane, 1968）。责任分散使得每个人所知觉到的主体责任会减小。人工智能对于人类来说是外群体，个体更重视人类合作伙伴而不是人工智能（Gombolay et al., 2015）。因此，当人与人工智能合作时，责任分散的程度可能会减少。也就是说个体在人机合作中比人-人合作时知觉到的主体责任更大。例如，Hinds 等人(2004)要求被试与合作伙伴收集组装各种物体所需的零件，实验中操纵了合作伙伴的属性和地位，被试需要评价他们感到自己

对任务绩效有多大责任。结果表明，与类似机器的机器人合作相比，个体在与类人的机器人合作时，知觉到对任务的责任更小。也就是说，当机器人更像人的时候，个体更愿意把完成任务的责任分给合作伙伴，知觉到的主体责任减小。

假设 2：个体知觉到的主体责任在人机合作（vs. 人-人）对个体风险决策影响中起中介作用。人机合作比“人-人”合作时，个体知觉到的主体责任更大，从而在风险决策中更冒险。

1.3 人机合作结果反馈与风险决策

人机合作中知觉到的主体责任与结果成败有关(Kim & Hinds, 2006; Lei & Rau, 2021)。以往研究发现人们将好的结果归于自己，而将坏的结果归于外部因素，即自我服务偏差（Miller & Ross, 1975）。据此，在成功结果反馈下，人们会认为自己对结果承担更大的责任。人工智能对于人类来说是外群体，与人工智能合作（vs. 人-人合作）时，个人可能会认为自己对结果承担的责任更大。以往关于人-机合作中积极结果反馈的研究也大都发现，人们将更多的责任归于自己，而不是人工智能。例如，You 和 Nie 等人 (2011)要求个体先重复机器人的动作，然后机器人会给予评价，如果获得积极反馈，个体归因于自己，如果获得消极反馈，个体则归因于机器人。Lei 和 Rau (2021)考察了人机合作中对不同地位机器人的责任归因，结果发现，积极结果反馈条件下，不管机器人的地位如何，个人都认为更多归因于自己。

以往关于消极结果反馈的研究结论却与自我服务偏差不完全一致（Franklin et al, 2022）。有的研究发现，人们将更多消极结果的责任归于人工智能。Kim 和 Hinds（2006）发现，当护士跟机器人一起合作，出现意外时护士将大部分的责任归因于机器人。被试会避免将很多失败责任归咎于与人相似的机器人，但会归于类机的机器人，类机器人可以减少用户的责任归因和内疚感 (Matsui & Koike, 2021) 。在人和机器人组成的群体中，当失败时，机器人成员比其他人类成员受到更多的谴责（Lei & Rau, 2021）。然而，另外一些研究则发现，出现消极结果时，人们对自己和人工智能承担的责任没有显著差异，甚至表现出对人工智能更多的宽容。例如，要求被试对人机共驾情境 AI 司机或人类司机的责任进行评分，结果表明积极结果下，人们对人工智能司机会有更多的赞扬，消极结果下对人类司机和机器人司机的责任归因无差异(Hong et al., 2021)。Awad 等人(2019)研究发现相比于只有一个驾驶员的情况下，当人机共享控制车辆的情况下犯错时，公众归咎于机器的责任会减少。根据自我归类理论，人工智能是不同于人类的外群体。即使结果失败，人机合作时（与人-人合作相比），个人可能会认为自己对结果承担的责任更大。

这样，虽然不管成功反馈还是失败反馈，“人-机”合作都会比“人-人”合作条件下个体知觉到的主体责任更大，但是成功反馈下，个体在人机合作时知觉到的责任会更大。因此，我们推测，结果反馈调节了知觉到的主体责任在“人-机”和“人-人”合作对个体风险决策影响中的中介作用。

假设 3：结果反馈调节了知觉到的主体责任在“人−机”和“人−人”合作对个体风险决策影响中的中介作用。具体来说，成功反馈比失败反馈时，个体在“人−机”合作比“人−人”合作时知觉到的主体责任更大，从而会更冒险。

2 实验 1 人−机合作对个体风险决策的影响

2.1 实验 1a

2.1.1 被试

从某大学共招募到 100 名大学生，剔除 11 名不相信实验真实性的被试，有效被试 89 名。其中男生 26 名，女生 63 名。平均年龄为 19.35 岁，标准差为 1.55 岁。使用 G*power 3.1 计算研究所需样本量(Faul et al., 2007)，被试内重复测量方差分析，效应量 $f = 0.25$，统计检验力 $1 - \beta = 0.80$，$a = 0.05$，测量次数 = 3，计算样本量为 28 人。

2.1.2 实验设计

采用单因素被试内设计，分为三个水平“人-机”合作，“人-人”合作、无合作伙伴（控制组），因变量为个体冒险水平，通过仿真气球冒险任务测量，以未打爆气球的平均打气次数来衡量。

2.1.3 实验程序

（1）合作情景的操作。被试来到实验室之后，被告知这是一项风险决策任务，接下来被试将通过抽签的方式确定完成风险决策任务的方式，选项有三种：与人合作、与机器人合作、独自完成（控制条件）。①“人−机”合作情境下，被试被告知：“你的合作伙伴是在隔壁机器人实验室的机器人（实验室门口张贴机器人实验室字样），你们将共同完成打气球任务，你们在实验中的收益和损失将会共同承担，我们会根据收益的多少按一定的比例为你们现场发放报酬，你的伙伴也会得到相应的报酬”。抽签结束后，被试在电脑上点击空格键开始实验。根据 Posard 和 Rinderknecht (2015)的做法，合作情境下屏幕上显示连接进度百分比的“旋转轮”图像（见图 1），下方配文字“请等待，我们正在与你的伙伴进行连接”，本研究中进度“旋转轮”显示了 10 秒，连接成功后屏幕显示：“连接成功！请为你们的团队起一个名字，并写下一句话为你们的团队加油”。之后，被试完成打气球任务。②“人−人”合作情景下，被试被告知合作伙伴是在对面实验室的人类伙伴（实验室门口张贴人类实验室）。之后，被试完成打气球任务。③控制组条件下屏幕中不出现进度“旋转轮”，指导语如下：“你将独自完成打气球任务，你将自己承担在实验中的收益和损失，我们会按照收益的多少按一定比例给你报酬”。通过抽签对三种合作情境的顺序进行了平衡。



图 1 合作情景的操作

（2）仿真气球冒险任务（BART, Balloon Analogue Risk Task）

在 BART 任务中，屏幕上会呈现一个模拟的气球，被试只需要按键，就能逐渐吹大这个气球，每次吹气球都有一定的收益，同时也伴随着气球爆破的风险，气球越大意味着获得的收益越大，但也会有更高的爆破风险，如果气球被吹爆，该气球的收益便为零。被试可以选择继续以追求更大的收益，也可以随时停止吹气球来保存现有的收益。该任务需要被试连续地做出决策，为了更大的收益继续冒险还是停止冒险以获得当前的收益。BART 任务中，“未爆气球的平均打气次数”是衡量风险偏好的主要行为指标，未打爆气球的平均打气次数越高，被试越偏好冒险。未爆气球平均打气次数＝未爆气球总打气次数 / 未爆气球的个数。为了避免气球因为随机打爆而出现打一下就爆炸的情况，本研究结合实际情况和实验目的，设置最小打爆次数为三下，本研究气球的打爆次数量程为 3~30 下，每充一次气球的收益为 0.5 元。本研究采用未爆气球的平均打气次数作为被试冒险水平的衡量指标，被试与合作伙伴一起完成打气球任务，最后按照团队共同受益，按一定比例折合现金奖励。

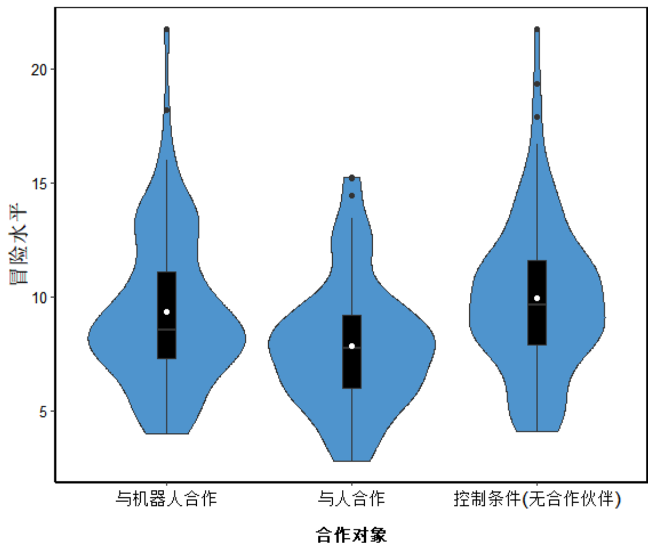
（3）检验合作情境是否操作成功。参照前人的操纵检验 (Posard & Rinderknecht, 2015)，在研究结束时询问被试：你认为你的合作伙伴是真实的人类或机器人、是否相信该实验情境的真实性、以及团队的名字。为保证数据的真实性，剔除11名不相信该实验情境的被试。

（4）实验结束后，对被试表示感谢，发放报酬。

2.1.4 结果分析

为了考察不同合作对象（与人合作、与机合作、控制组）对个体风险决策的影响，重复测量方差分析结果显示：F (2, 176) = 33.44，p < 0.01，η_p^2 = 0.28，进一步事后检验显示，与人合作时个体冒险水平最低 (M = 7.84, SD = 2.79)，显著低于与机器人合作时的冒险水平 (M = 9.33, SD = 3.36)，t(88)= -6.94, p < 0.001, d = 0.74；与机器人合作时的冒险水平显著低于控制组的冒险水平(M = 9.92, SD = 3.43)，t(88)= -2.51, p = 0.014, d = 0.27；与人合作时的冒险水平也显著低

于控制条件下的冒险水平，$t(88)= -6.43$, $p < 0.001$, $d = 0.68$。三种合作情景下个体冒险水平的描述统计见图 2。



注：中心白点表示平均数，黑色的方块表示四分位数的范围

图 2 实验 1a 不同实验条件下的冒险水平

2.2 实验 1b

实验 1a 中，被试与合作伙伴之间并没有互动。实验 1b 进一步在互动情景下考察人机合作对个体风险决策的影响。

2.2.1 被试

从某大学招募参加者 151 人，其中男生 39 名，女生 112 名。平均年龄为 20.99 岁，标准差为 2.45 岁。使用 G*power 3.1 计算研究所需样本量(Faul et al., 2007)，效应量 $d = 0.50$, 统计检验力 $1-\beta = 0.80$, $a = 0.05$ 时，双侧独立样本 t 检验需要的样本量为 128 人。

2.2.2 实验设计

采用单因素被试间设计，分为两个水平，即“人-机”合作，“人-人”合作，因变量为个体冒险水平，通过仿真气球冒险任务测量，以未打爆气球的平均打气次数来衡量。

2.2.3 实验程序

(1) 合作情景的操作。被试来到实验室之后，被告知这是一项在线互动的风险决策游戏，接下来被试将被随机分到两种不同的合作情境：与人合作、与机器人合作。“人–机”合作情境下，被试被告知:“您好，本轮游戏你的同伴是人工智能机器人 Alpha03， 接下来你将和他共同完成本轮的打气球游戏，你们同步完成游戏，互相可以看见对方的选择。你可以按“F”键充气，按“J”键停止打气，每充一次气（即按“F”键一次）会获得 0.5 元的奖励，气球在充气的过程中随时会爆炸，如果气球爆炸，你们在该气球上的当前收益会被清零，随后进入下

一轮打气球。在气球未爆炸前的任意一个时刻，你均可以通过按“J”键停止充气，以保存当前气球的收益，并进入下一轮实验。整个过程将呈现 30 个气球，最后我们将把你和同伴各自在游戏中的收益合并在一起求平均，按一定的比例折算成现金奖励发放给你们。如果已理解任务要求，请按空格键开始实验”。根据 Posard 和 Rinderknecht (2015)的做法，合作情境下屏幕上显示连接进度百分比的“旋转轮”图像（见图 1），本研究中进度“旋转轮”显示了 10 秒，连接完成后，屏幕上呈现待充气的气球，被试开始游戏。②“人-人”合作情景下，被试被告知合作伙伴是系统随机匹配的人类同伴，其他与“人-机”合作条件下相同。实验界面如下图所示。



图 3 人机互动情景下的风险决策

(2) 仿真气球冒险任务(BART, Balloon Analogue Risk Task)

仿真气球冒险任务同实验 1a。该任务需要被试连续地做出决策：为了更大的收益继续冒险还是停止冒险以获得当前的收益。屏幕左边呈现被试的选择（停止打气或继续打气），屏幕右边呈现合作伙伴的选择（停止打气或继续打气）。合作伙伴的选择是由程序设定好的随机选择。每一轮打气球游戏中，被试先进行停止或打气的选择，等被试做出选择后，屏幕右半部分立即呈现合作伙伴的选择。

(3) 检验合作情境是否操作成功。为了对合作进行操纵检验，要求被试回答“在实验任务中，你的合作同伴是机器人还是人类同伴”。

(4) 操作检验后，被试完成正式游戏，并填写人口学信息问卷。实验结束后，对被试表示感谢，发放实验报酬。

2.2.4 结果分析

独立样本 t 检验结果表明，“人-机”合作条件下个体的冒险水平(M = 10.77, SD = 3.23)显著高于“人-人”合作条件(M = 8.99, SD = 3.49)，t(149)= 3.24，p = 0.001, d = 0.53。

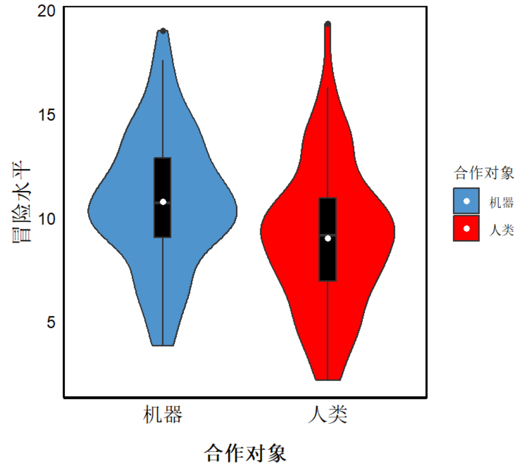


图 4 不同合作对象的冒险水平

2.3 实验 1 讨论

实验 1a 结果发现，个体单独决策时冒险水平最高，显著高于与人合作、与机器人合作，与研究假设一致。这说明在合作情景中，不管是与人工智能合作还是与人类合作，个体的冒险水平都会比自己单独决策时更低。可能的原因是，在合作情景下，个体的决策不仅影响到自己，还会影响到合作伙伴，万一失败了会使合作伙伴的收益减少，从而冒险水平降低。实验 1a 和 1b 结果均发现，个体在与人工智能合作时的冒险水平显著高于与人合作时，与研究假设一致。这说明，同样是合作情景下，与人工智能合作时，个体的冒险水平高于与人合作时，可能的原因是，机器人不是人类，对于人类来说是外群体，个体在与机器人合作时，较少考虑对合作伙伴的责任，从而冒险水平相对较高。以往的研究表明，伙伴亲疏关系影响会风险决策的结果，有研究对比了独自决策、亲密情侣在场、普通人在场时的个体冒险水平，结果发现亲密情侣在场使个体最保守，其次是与普通人共同决策，独自决策最冒险(Silva et al., 2020)。

3. 实验 2 知觉到的主体责任在人机合作对个体风险决策影响中的作用

实验 2 将进一步考察人机合作影响个体冒险行为的中介机制。

3.1 被试

在某大学招募被试 199 人，其中男生 47 人；与人合作组 105 人，男生 22 人，与机器人合作组 94 人，男生 25 人，平均年龄为 19.71 岁，标准差为 2.45 岁。借鉴 Schoemann 等人(2017)探究中介效应所需样本量的计算方法，5000 次抽样，α，β 和 c 路径大小设置为 0.35、0.25、0.1（中等大小），结果发现，至少需要 145 名被试，才能使统计检验力达到 80%。

3.2 实验设计

采用被试间实验设计，自变量为合作对象（“人-人”合作、“人-机”合作），中介变量为知觉到的主体责任，因变量为个体的冒险水平，通过仿真气球冒险任务测量，以未打爆气球的平均打气次数为指标。

3.3 实验流程

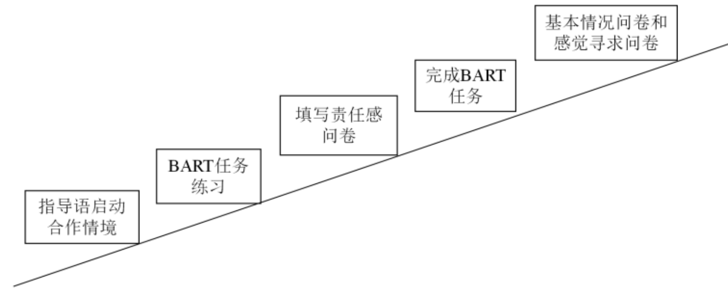


图 5 实验 2 实验流程图

（1）合作情景操纵

“人-机”合作和“人-人”合作情境操作方法同实验 1a。

（2）知觉到的主体责任测量

采用 Hinds 等人（2004）编制的“人-机”合作中的责任感问卷来测量知觉到的主体责任，该量表共 5 道题，例如，“你认为出色的完成这项任务是你的责任”“你认为这项任务是你的事情”。1~7 点评分，得分越高，被试知觉到在该任务中承担的责任越大，反之则越小。三名心理学研究生对原始英文版问卷进行翻译，综合讨论后确定中文版问卷，并由相关领域专家回译。对中文版量表进行单因素验证性因素分析，拟合指标如下：$\chi 2/df = 2.214$，RMSEA = 0.078， NFI = 0.969，RFI = 0.922， IFI = 0.983，TLI = 0.956，CFI = 0.982，GFI = 0.983。本研究中该量表的内部一致性信度 Cronbach’s $\alpha = 0.731$，符合心理测量学要求。

（3）风险决策

实验采用仿真气球冒险任务，实验程序同实验 1。

（4）感觉寻求测量

为了对被试的风险偏好特质进行控制，我们采用赵闪(2004)编制的中国版大学生感觉寻求问卷。该问卷由两个因子构成：兴奋和冒险寻求因子、去抑制因子，共包含 36 个项目，例如“爬陡峭的山”。

3.4 结果分析

3.4.1 不同合作情景下个体的冒险水平和知觉到的主体责任

独立样本 t 检验的结果显示，与机器人合作(M=11.41, SD=3.24)比与人合作(M=8.55, SD=2.55)冒险水平更高，$t(197)=6.69$, $p<0.001$, $d=0.95$。与机器人合作(M=5.06, SD=0.57)比与人

合作(M=4.15, SD=0.37)时个体知觉到的主体责任更大，$t(197)=13.41$, $p<0.001$, d=1.88。不同合作情景下个体冒险水平和知觉到的主体责任描述统计见图 6。

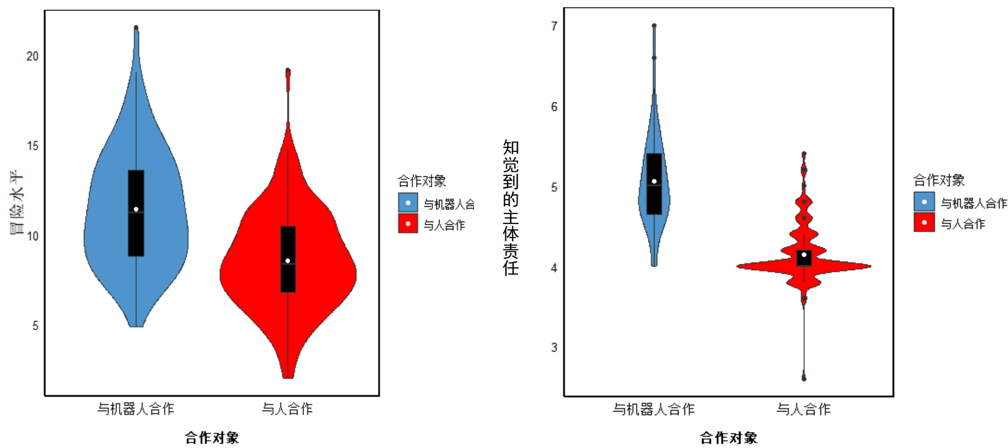


图 6 不同合作情境下个体冒险水平和知觉到的主体责任

3.4.2 知觉到主体责任的中介作用分析

为了考察知觉到的主体责任在合作情景对个体冒险水平的影响中的中介作用，我们首先进行了三步回归分析，结果见表 1，从方程 1 可以看到，合作对象可以显著正向预测个体冒险水平(β=0.42, t=6.52, $p<0.001$)。从方程 2 可以看到，合作对象显著正向预测知觉到的主体责任(β=0.69, t=13.32, $p<0.001$)。从方程 3 可以看到，加入知觉到的主体责任以后，知觉到的主体责任显著正向预测个体冒险水平(β=0.23, t=2.59, p=0.01)，合作对象对个体冒险水平的直接效应依然显著(β=0.26, t=3.00, p=0.003)。因此，知觉到的主体责任在合作对象对个体冒险水平的影响中起部分中介作用。Bootstrap 中介效应检验发现，中介效应值为 0.1572，95%的置信区间上不包括 0，95%CI=[0.0573, 0.2849]，中介效应显著，中介效应占总效应的占比为 37.36%。

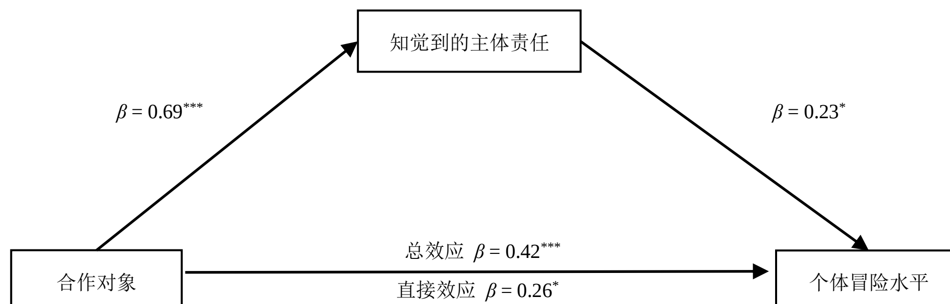
表 1 合作对象、知觉到的主体责任对个体冒险水平影响的回归分析

预测变量	方程 1（因变量：平均打气次数）			方程 2（因变量：知觉到的主体责任）			方程 3（因变量：平均打气次数）		
	β	*SE*	*t*	*β*	*SE*	*t*	*β*	*SE*	*t*
常数		2.37	2.93**		0.38	8.98***		2.78	1.09
年龄	-0.002	0.09	-0.03	0.03	0.01	0.55	-0.01	0.09	-0.13
性别	-0.16	0.50	-2.44*	-0.12	0.08	-2.38*	-0.13	0.50	-2.00*
感觉寻求	-0.01	0.03	-0.12	-0.03	0.004	-0.62	<0.001	0.03	-0.01
合作对象	0.42	0.43	6.52***	0.69	0.07	13.32***	0.26	0.59	3.00**
知觉到的主体责任							0.23	0.45	2.59*
F	F (4, 194) =12.86***			F (4, 194) = 47.16***			F (5, 193) =11.93***		
R^2	0.21			0.49			0.24		

注：合作对象为虚拟变量，1=人工智能， 0=人。控制变量为年龄、性别、感觉寻求。

表 2 知觉到主体责任的中介作用

	effect	BootSE	BootCI 下限	BootCI 上限	a*b/c
知觉到的主体责任	0.1572	0.0576	0.0573	0.2849	37.36%



注：合作对象为虚拟变量，1=人-机合作，0=人-人合作

图 7 知觉到主体责任的中介作用

3. 5 实验 2 讨论

实验 2 结果发现，知觉到的主体责任在“人-机”合作（vs.“人-人”合作）对个体风险决策影响中起部分中介作用，与“人-人”合作相比，个体在人机合作中知觉到的主体责任更

大，认为合作任务是个人的义务和责任，因此更冒险，支持了假设 2。前人研究也发现，与机器人合作时个体承担较多的责任(Hinds et al., 2004)。以往关于责任感对风险决策的影响研究发现，责任感越大，风险决策越保守。例如，Wang 等人(2018)的研究中，一种条件下，被试单独作风险决策，另外一种条件下，被试代表他/她所在的三人群体作风险决策，结果发现，对群体成员的责任感会使代表群体决策比个体决策更保守。Bolton 等人(2015)的研究中，一种条件是被试单独作风险决策，只影响自己的收益，另一种条件是被试为自己和另外一名伙伴作风险决策（决策不仅影响自己的收益还会影响到伙伴的收益），结果发现，在后一种条件下，对伙伴的责任感使个体更保守。本研究发现，个体在合作中知觉到的主体责任更大，风险决策中更冒险，与以往研究结果不一致。这与责任感的定义不同有关。本研究中的责任感是个体知觉到的主体责任，即在合作中自己所分担对于完成任务的责任，而不是指为伙伴决策的社会关系责任。在合作中个体知觉到的主体责任越大，更倾向认为合作任务是个体自己的责任，因此更倾向冒险，而为伙伴做决策的社会关系责任感越大，更会担心影响其他人的收益，因此个体在风险决策中越谨慎。

4．实验 3 结果反馈在人机合作对风险决策影响中的作用

在实验 2 的基础上，实验 3 进一步考察了结果反馈在人机合作对风险决策影响中的调节作用。

4.1 被试

在某大学招募被试 199 人，其中男生 47 人；与人合作组 105 人，男生 22 人，与机器人合作组 94 人，男生 25 人，平均年龄为 19.71 岁，标准差为 2.45 岁。使用 G*power 3.1 计算研究所需样本量(Faul et al., 2007)，在两因素被试间设计的方差分析中, 效应量 f = 0.25，统计检验力 $1-\beta = 0.80$， $a = 0.05$，计算需要样本量为 128 人。

4.2 实验设计

采用 2(合作对象：人−机、人−人)× 2（结果反馈：成功反馈、失败反馈）被试间实验设计，中介变量为知觉到的主体责任，因变量为个体的冒险水平，通过仿真气球冒险任务测量，以未打爆气球的平均打气次数为衡量个体冒险水平的指标。

4.3 实验流程

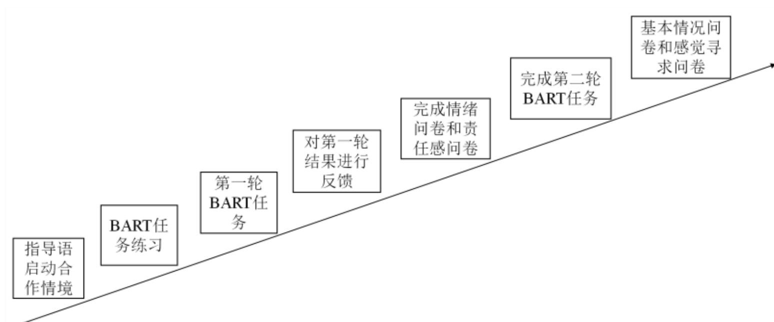


图 8 实验 3 流程图

（1）合作情景操纵

自变量合作情景操纵方法同实验 1a。

（2）结果反馈

在实验 2 的基础上，对第一轮任务的结果进行反馈，将团队的平均成绩与所有参加该实验团队的平均成绩对比，作为结果反馈的依据。高于平均成绩为成功反馈，在第一轮收益的基础上获得额外的报酬奖励；低于平均成绩为失败反馈，在第一轮团队收益的基础上扣除一部分收益作为惩罚。

（3）知觉到的主体责任测量

知觉到的主体责任问卷改编自 Lei 和 Rau(2021) 人机合作中的责任归因问卷，采用其中测量对自己责任归因的 5 道题目。例如"在这项任务中我贡献了大部分好的/坏的表现""我在多大程度上要对团队的成功/失败负责？"问卷为 1~7 点评分，得分越高，被试在该任务中知觉到的主体责任越大，反之则越小。该量表已经应用于中国大学生的研究(Lei & Rau, 2021)。该问卷在本研究中的 Cronbach's α = 0.957。

为了控制结果反馈后被试情绪对风险决策的影响，结果成败反馈之后被试填写正负情绪问卷（PANAS）(Watson et al., 1988；黄丽等，2003)。该量表由正性情绪分量表 10 道题和负性情绪分量表 10 道题两部分组成，共 20 道题，如："感兴趣的、心烦的"，1~5 评分，分别计算积极情绪和消极情绪得分。

（4）仿真气球冒险任务

结果反馈后，被试完成仿真气球冒险任务同实验 1，以未打爆气球的平均打气次数衡量个体冒险水平。

（5）为控制个体特质冒险倾向的影响，BART 任务结束后填写感觉寻求问卷，同实验 2。

4.4 结果分析

4.4.1 不同结果反馈、合作对象对个体冒险水平的影响

以个体冒险水平为因变量，将个体的性别、年龄、感觉寻求、积极和消极情绪作为协变量，2（合作对象：人、机）× 2（结果反馈：成功、失败）完全随机方差分析结果显示，合作对象的主效应显著，$F(1, 190)=25.55$, $p<0.001$, $\eta_p^2=0.12$，与机器人合作时个体的冒险水平($M=11.82$, $SD=3.23$)显著高于与人合作($M=9.31$, $SD=3.04$)；结果反馈的主效应不显著，$F(1, 190)=0.03$, $p=0.871$，成功反馈下的个体冒险水平($M=10.87$, $SD=2.98$)和失败反馈下的个体冒险水平($M=10.12$, $SD=3.70$)无显著差异。合作对象和结果反馈的交互作用不显著，$F(1, 190)=0.62$, $p=0.432$。成功反馈时，与机器人合作时个体冒险水平($M = 11.88$, $SD = 2.79$)显著高于与人合作时的个体冒险水平($M = 9.85$，$SD = 2.83$)，$t(98)= -3.60$, $p=0.001$, $d=0.72$。当失败反馈时，与机器人合作时个体冒险水平($M = 11.76$, $SD = 3.70$)同样显著高于与人合作时的个体冒险水平($M = 8.81$，$SD = 3.16$)，$t(97)= -4.28$，$p<0.001$, $d=0.85$（见图 9）。

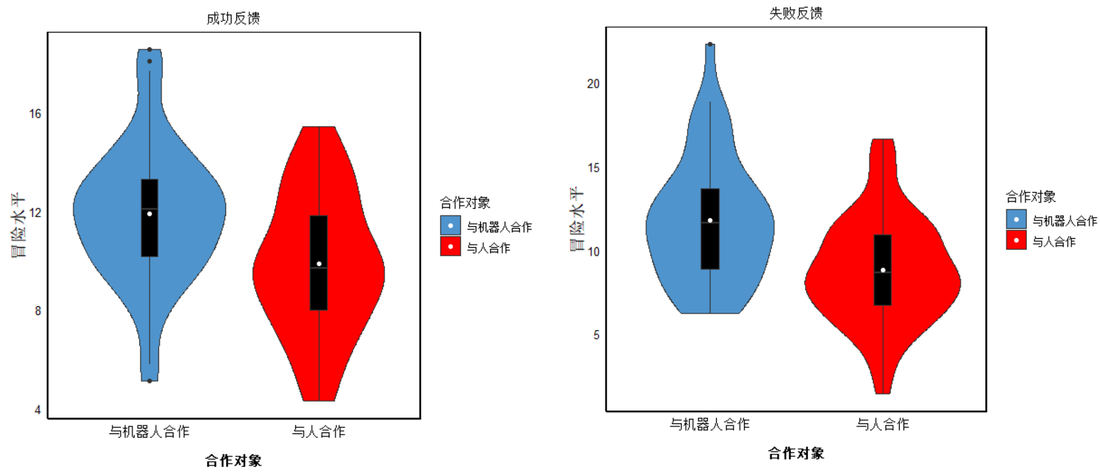


图 9 不同结果反馈与合作情景下的冒险水平

4.4.2 不同结果反馈和合作对象对知觉到的主体责任的影响

以知觉到的主体责任为因变量，将个体的性别、年龄、感觉寻求、积极和消极情绪作为协变量，2（合作对象：人、机）× 2（结果反馈：成功、失败）完全随机方差分析结果显示，合作对象的主效应显著，$F(1, 190)=10.69$, $p=0.001$, $\eta_p^2=0.05$，与机器人合作时知觉到的主体责任($M = 4.72$, $SD = 0.98$)显著高于与人合作时的知觉到的主体责任($M = 4.32$, $SD = 0.74$)；结果反馈的主效应显著，$F(1, 190)=18.01$, $p<0.001$, $\eta_p^2=0.09$，失败反馈时知觉到的主体责任($M = 4.74$, $SD = 0.82$)显著高于成功反馈时($M = 4.29$, $SD = 0.89$)；合作对象和结果反馈的交互作用显著，$F(1, 190)=6.82$, $p=0.010$, $\eta_p^2=0.04$。从图 10 可以看到，成功反馈时，与机器人合作时知觉到的主体责任（$M = 4.64$ $SD = 1.00$）高于与人合作（$M = 3.93$, $SD = 0.57$），$t(98)= -4.38$，$p<0.001$, $d=0.86$；失败反馈时，与机器人合作时知觉到的主体责任($M = 4.81$ $SD = 0.93$)与“人-人”合作

时(M = 4.68, SD = 0.71)无显著差异, $t(97)$= -0.77, p=0.441。

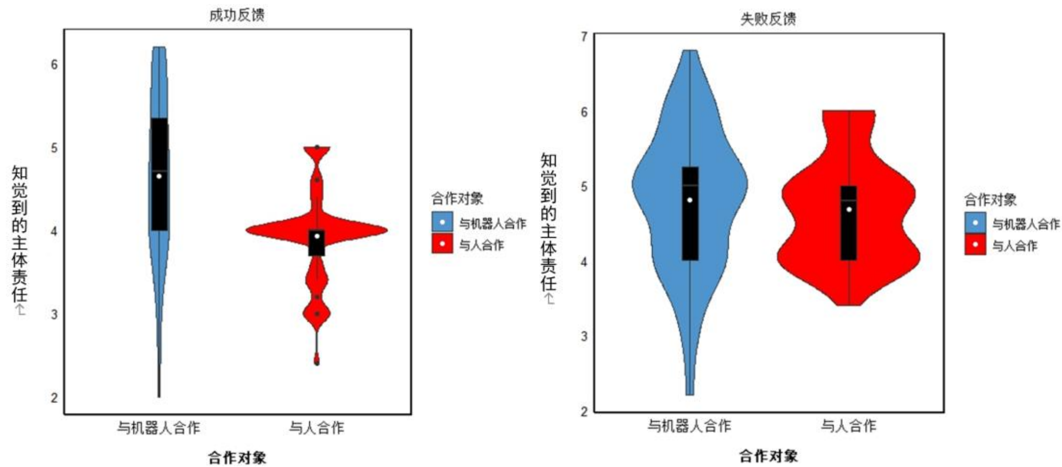


图 10 不同结果反馈与合作对象下知觉到的主体责任

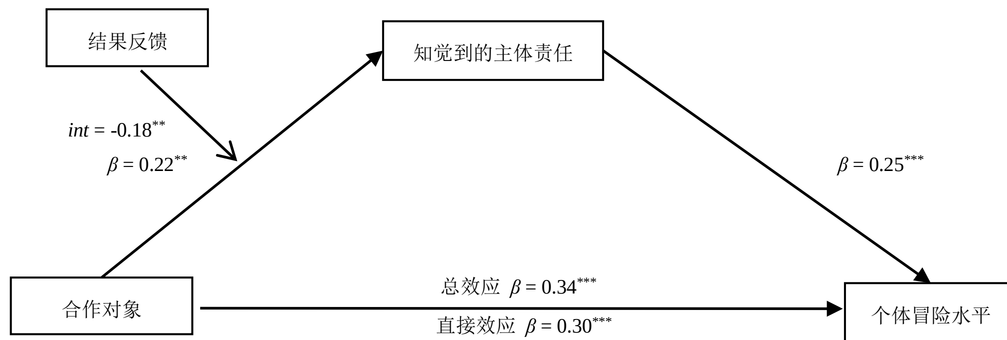
4. 4. 3 结果反馈的调节作用分析

根据温忠麟和叶宝娟(2014)的观点，检验有调节的中介模型需要对三个回归方程的参数进行检验：（1）方程 1 估计调节变量（结果反馈）对自变量（合作对象）与因变量（冒险水平）之间关系的调节作用；（2）方程 2 估计调节变量（结果反馈）对自变量（合作对象）与中介变量（知觉到的主体责任）之间关系的调节效应；（3）方程 3 估计调节变量（结果反馈）对中介变量（知觉到的主体责任）与因变量（冒险水平）之间关系的调节效应以及自变量（合作对象）对因变量（冒险水平）残余效应的调节效应。结果表明，结果反馈与合作对象的乘积项对知觉到的主体责任作用显著(β = -0.18, t =-2.61, p = 0.01)，对个体冒险水平的预测作用不显著(β = 0.06, t = 0.94, p = 0.35)，结果反馈和知觉到的主体责任的交互项对冒险水平的预测作用不显著(β = 0.13, t = 0.07, p = 0.07)，说明结果反馈在合作对象对知觉到的主体责任影响中起调节作用（路径的前半部分）。

表 3 有调节的中介效应检验 (N = 199)

预测变量	方程 1（因变量：冒险水平）			方程 2（因变量：知觉到主体责任）			方程 3（因变量：冒险水平）		
	β	SE	t	β	SE	t	β	SE	t
常数	0.003	0.07	0.04	-0.01	0.07	-0.15	-0.03	0.07	-0.44
年龄	-0.01	0.07	-0.12	0.01	0.07	0.20	-0.01	0.06	-0.17
性别	-0.10	0.07	-1.47	-0.07	0.07	-1.05	-0.08	0.07	-1.21
感觉寻求	0.06	0.07	0.84	-0.01	0.07	-0.15	0.07	0.06	0.03
积极情绪	0.14	0.07	1.86	0.14	0.07	1.87	0.10	0.07	1.36
消极情绪	-0.09	0.07	-1.22	-0.05	0.07	-0.70	-0.09	0.07	-1.31
合作对象	0.34	0.07	5.05***	0.22	0.07	3.28**	0.30	0.07	4.50***
结果反馈	-0.02	0.08	-0.20	0.33	0.08	4.36***	-0.09	0.08	-1.20
知觉到的主体责任							0.25	0.07	3.57***
结果反馈×合作对象	0.05	0.07	0.79	-0.18	0.07	-2.61**	0.06	0.07	0.94
结果反馈×知觉到的主体责任							0.13	0.07	1.82
F	F (8,190) = 5.67***			F (8,190) = 5.13***			F (10,188) = 6.50***		
R^2	0.19			0.18			0.26		

注：控制变量为年龄、性别、感觉寻求、积极情绪、消极情绪。均为标准化数据。



注：合作对象为虚拟变量，1=人-机合作，0=人-人合作。

图 11 有调节的中介作用

通过 Bootstrap 法进行有调节的中介效应检验(Hayes, 2018)。选择模型 7，以合作对象为自变量（与人合作 = 0，与机器人合作 = 1），个体冒险水平为因变量，知觉到的主体责任为中

介变量，结果反馈(成功反馈=0，失败反馈=1)为调节变量，将性别、感觉寻求、积极情绪、消极情绪作为控制变量。结果显示（见表 4），成功反馈下，知觉到的主体责任的中介作用显著 95%CI [0.0246, 0.1622]，失败反馈下，知觉到的主体责任的中介作用不显著 95%CI [-0.0329, 0.0607]，结果反馈调节了知觉到的主体责任的中介作用。

表 4　成功/失败反馈下知觉到的主体责任的中介效应

中介变量	结果反馈	effect	BootSE	BootCI 下限	BootCI 上限
知觉到的主体责任	成功（0）	0.0821	0.0354	0.0246	0.1622
	失败（1）	0.0094	0.0233	-0.0329	0.0607

简单斜率分析表明，在成功反馈下，“人−人”合作和“人−机”合作个体知觉到的主体责任有显著差异($\beta = 0.40$, $t = 4.34$, $p < 0.001$)， 与“人−人”合作条件相比，人机合作条件下，个体知觉到的主体责任更大。失败反馈下，“人−人”合作和“人−机”合作个体知觉到的主体责任无显著差异 ($\beta = 0.08$, $t = 0.78$, $p = 0.43$)。在失败反馈下，无论合作伙伴是机器人还是人，被试都倾向于自己为团队的失败承担更多的责任。

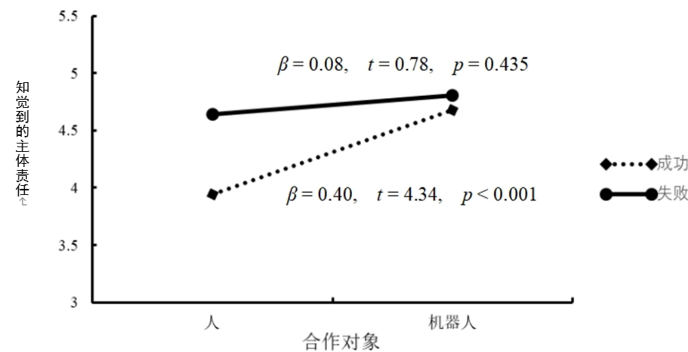


图 12 结果反馈的调节作用

4.5 实验 3 讨论

实验 3 发现，结果反馈调节了知觉到的主体责任在“人−机”和“人−人”合作对个体风险决策影响中的中介作用。成功反馈下，与“人−人”合作相比，人机合作条件下，个体知觉到的主体责任大，知觉到的主体责任在人机合作对个体风险决策影响中起中介作用；失败反馈下，“人−人”合作与“人−机”合作条件下，个体知觉到的主体责任没有显著差异，知觉到的主体责任在人机合作对个体风险决策影响中的中介作用不显著，部分支持了假设 3。本研究发现，在成功反馈下，人机合作时，个体承担的责任更强，这与自我服务归因偏差（Miller &

Ross，1975）一致，也与以往人机互动中发现的结果一致（Kim & Hinds，2006；You et al., 2011; Matsui & Koike, 2021; Lei & Rau, 2021)。然而，本研究发现，在失败反馈下，“人−人”合作与“人−机”合作条件下，个体知觉到的主体责任没有显著差异。失败反馈下，无论与人合作还是与机器人合作，个体自己均承担较多的责任，可能与个体对机器人的包容有关（毕竟机器人不是人类)，也可能与个体对自己的信心有关。例如，Chong 等人(2022)的研究发现，是对自己的信心，而不是对 AI 的信心影响了被试对 AI 建议的接受，当被试接受人工智能的建议并得到负面反馈的时候，他们会将 AI 的责任归咎于自己，责怪自己未能检测到 AI 错误。

5 总讨论

本研究从人机合作中主体责任的视角出发，通过四个实验，考察了人机合作对个体风险决策的影响及其机制，结果发现：第一，不管与人合作，还是与人工智能合作，都会比单独做决策时更谨慎；“人-机”合作比“人-人”合作时，个体会更冒险。第二，“人-机”合作比“人-人”合作时更冒险，部分通过了个体在合作中知觉到的主体责任的中介作用。具体说，个体在“人-机”合作中比“人-人”合作时，知觉到更多的个人主体责任，更倾向于把合作任务看作是自己的责任，因此更冒险。第三，结果反馈调节了知觉到的主体责任在人机合作对个体风险决策影响中的中介作用。成功反馈下，与“人−人”合作相比，人机合作时，个体承担更多责任，知觉到的主体责任在人机合作对个体风险决策影响中的中介作用显著；失败反馈下，“人−人”合作与“人−机”合作相比，个体知觉到的主体责任没有显著差异，知觉到的主体责任在人机合作对个体风险决策影响中的中介作用不显著。

5.1 人机合作使个体更倾向冒险

实验 1a 和 1b 结果均发现，个体在与人工智能合作时的冒险水平显著高于与人合作时，与研究假设一致。可能的原因是，机器人对于人类来说是外群体，个体在与机器人合作时，较少考虑对合作伙伴的责任，从而冒险水平相对较高。以往研究表明，伙伴关系会影响风险决策的结果。有研究对比独自决策、亲密情侣在场、普通人在场时的个体冒险水平，结果发现亲密情侣在场使个体最保守，其次是与普通人共同决策，独自决策最冒险(Silva et al., 2020)。

5.2 知觉到的主体责任与结果反馈在人机合作对个体风险决策影响中的作用

实验 2 结果发现，知觉到的主体责任在“人-机”合作和“人-人”合作对个体风险决策影响中起部分中介作用，与“人-人”合作相比，在人机合作中个体知觉到主体责任更大，因此更冒险，支持了假设 2。以往关于责任感对风险决策的影响研究发现，责任感越大，风险决策越保守。例如，Wang 等人(2018)的研究中，一种条件下，被试单独做决策，另外一种条件下，被试代表他/她所在的三人群体做决策，结果发现，对群体成员的责任感会使为群体决策比个体决策更保守。Bolton 等人(2015)的研究中，一种条件是被试单独做决策，另一种条件是

被试为自己和另外一名伙伴做风险决策，结果发现，在后一种条件下，对伙伴的责任感使个体更保守。我们认为不同的研究结果与对责任感的定义不同有关。本研究基于社会认知中主体能动性—社会关系性双重视角模型的研究（Abele & Wojciszke，2014），区分了两类责任感，即主体责任和社会关系责任。主体责任指在合作中个体自己对于目标实现和完成任务所承担的责任；社会关系责任指在个体在合作中为他人所承担的责任。本研究从主体责任的视角，考察人机合作对风险决策的影响，而以往研究大多是从社会关系责任的视角出发。在合作中个体知觉到的主体责任越大，更倾向认为合作任务是个体自己的责任，因此更倾向冒险，而为伙伴做决策的社会关系责任感越大，更担心影响其他人的收益，因此个体在风险决策中更保守。

实验 3 发现结果反馈调节了知觉到的主体责任在人机合作对个体风险决策影响中的中介作用。成功反馈下，与“人-人”合作相比，人机合作时个体知觉到的主体责任大，知觉到的主体责任在人机合作对个体风险决策影响中的中介作用显著。这与自我服务归因偏差(Miller & Ross，1975)一致，也与以往人机互动中的结果一致(Kim & Hinds, 2006; You et al., 2011; Matsui & Koike, 2021; Lei & Rau, 2021)。本研究发现，失败反馈下“人-人”合作与“人-机”合作条件下，个体知觉到的主体责任没有显著差异，无论与人合作还是与机器人合作，自己均承担较多的责任，知觉到的主体责任在人机合作对个体风险决策影响中的中介作用不显著。这可能与个体对机器人的包容有关（毕竟机器人不是人类），也可能与个体对自己的信心有关。例如，Chong 等人(2022)的研究发现，当被试接受人工智能的建议并得到负面反馈的时候，他们会将 AI 的责任归咎于自己，责怪自己未能检测到 AI 错误。另外，失败反馈下无论与人合作还是与机器人合作，自己均承担较多的责任，可能与社会赞许效应有关(巩金培 等, 2021)，失败反馈下这样做可能是为了得到更多好的评价；其次，可能与金额的奖励幅度有关，本研究采用仿真气球冒险任务，属于金额较小的金钱领域的风险决策，未涉及大数额的奖励，研究表明个体的风险偏好水平会随着奖赏幅度的增加而减小(Xu et al., 2018; 徐四华 等, 2013)。

5.3 理论贡献与实践价值

第一，在当前人工智能与人类共生的时代，人机合作使人更冒险还是更保守是值得研究的课题。目前关于人机合作的研究主要聚焦于合作时，人们对人工智能的感知、接受度、满意度等，例如，有学者发现，与机器人作为竞争对手时相比，机器人作为合作伙伴时加强了人对机器人的能力感知 (Oliveira, Arriaga, Correia, & Paiva, 2020)。然而，较少考察人机合作中人工智能如何影响人的风险行为。本研究发现，比起与人合作，人机合作时个体更冒险，并且个体知觉到的主体责任起到部分中介，拓展了人机合作的研究领域。

第二，当前关于责任感与风险决策的研究结论不一致，有的研究发现，责任感越强，风险决策越保守（Atanasov & Kunreuther, 2016; Bolton, et al., 2015; Charness & Jackson,2009; Wang, et al., 2018）；有的研究发现，责任感越强，风险决策越冒险（Hinds et al., 2014）。我们认为，

这与责任感的定义不一致有关。基于社会认知中主体能动性—社会关系性双重视角模型(Abele & Wojciszke, 2014)，本研究区分了两类责任感，即主体责任和社会关系责任。主体责任指在合作中个体自己对于目标实现和完成任务所承担的责任；社会关系责任指在合作中个体为他人所承担的责任。在合作中个体知觉到的主体责任越大，更倾向认为合作任务是个体自己的责任，因此更倾向冒险，而社会关系责任感越大，更会担心影响其他人的收益，因此个体在风险决策中越倾向风险规避。本研究从主体责任的视角，考察人机合作对风险决策的影响，不同于以往研究大多从社会关系责任的视角出发(Wang et al., 2018; Bolton et al., 2015)。本研究有助于深入理解不同决策情景下人机合作中个体的责任感，有助于解释当前关于责任感与风险决策的不一致研究结果。

本研究还具有重要的实践价值。未来人工智能会越来越多地参与到各种决策中，与人类共同决策，联合行动。研究人机合作对人们风险决策的影响，可以帮助人类更好地了解人机合作对人类冒险行为的影响，在人机合作时做出更优决策。

5.4 不足与展望

本研究首次从主体责任的视角考察人机合作对个体风险决策的影响，揭示了人机合作对个体风险决策的影响及其机制。然而，本研究还存在一些不足之处，值得未来进一步研究。首先，本研究人机合作是在实验室中创设情境实现的。实际生活中，当人与人工智能相处时间更长，如司机与自动驾驶机器人长时间相处，医生长期使用人工智能诊断设备。长时间地人机合作会人们的冒险行为产生怎样的影响呢？未来可以采用更真实的情景进一步考察人机合作对个体风险决策的影响。

第二，人机合作影响风险决策的机制需要进一步澄清。本研究发现个体在合作中知觉到的主体责任在人机合作对风险决策的影响中起到了部分中介作用。除此之外，人机合作影响个体风险决策的机制还有哪些？值得未来进一步研究。

第三，人机合作影响风险决策的边界条件还需要进一步深入研究。例如，人工智能决策的准确率可能会影响人机合作中个体的风险决策。未来需要进一步研究人工智能决策的准确率在人机合作风险决策中的作用。另外，研究表明风险决策存在领域特异性，如健康-安全领域、道德领域、社交领域等(岳灵紫　等, 2018)。本实验任务仅涉及金钱风险决策领域。一些人可能在金融投资上偏好高风险以追求高回报，而在健康决策领域则可能更加保守，倾向于选择风险较低的选项。因而不同领域中人机合作对个体风险决策的影响也可能会不同。未来研究可以运用更多样的决策任务探索不同领域中人机合作风险决策。最后，本研究中的被试总体上男生占比相对较少，未来需要进一步扩大男生样本。

6 结论

本研究通过四个实验得到以下结论：（1）不管与人合作，还是与人工智能合作，个体都会比单独做决策时更保守；“人-机”合作比“人-人”合作时，个体更冒险。（2）个体知觉到的主体责任部分中介了人机合作对风险决策的影响，即个体在“人-机”合作中比“人-人”合作时，知觉到更多的主体责任，从而更冒险。（3）结果反馈调节了知觉到的主体责任在人机合作对个体风险决策影响中的中介作用。成功反馈下，“人−机”合作比“人−人”合作，个体知觉到的主体责任更多，知觉到的主体责任在人机合作对个体风险决策影响中起中介作用；失败反馈下，“人−人”合作与“人−机”合作时个体知觉到的主体责任没有显著差异，知觉到的主体责任在人机合作对个体风险决策影响中的中介作用不显著。

致谢：感谢展讯科技（杭州）有限公司李春工程师在实验程序编制中的贡献。感谢杭州师范大学心理系赵亚婷、曾妍、彭彦琳、崔洁、孙亚茹，以及浙江大学心理与行为科学系王士祺在数据收集中的贡献。

参考文献

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *In Advances in Experimental Social Psychology* (Vol. 50, pp. 195-255). Academic Press.

Atanasov, P., & Kunreuther, H. (2016). Cautious defection: Group representatives cooperate and risk less than individuals. *Journal of Behavioral Decision Making, 29*(4), 372–380.

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2019). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour*, *4*(2), 134–143.

Bolton, G. E., Ockenfels, A., & Stauf, J. (2015). Social responsibility promotes conservative risk behavior. *European Economic Review, 74*, 109–127.

Charness, G., & Jackson, M. O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization, 69*(3), 241–247.

Chong, L., Zhang, G., Goucher-Lambert., K., Kotovsky., K., & Cagan, J. (2022). Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior, 127*, 107018.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8*(4, Pt.1), 377–383.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an f MRI investigation. *Neuroimage, 23*(2), 744–751.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Forgiarini, M., Gallucci, M., & Maravita, A. (2011). Racism and the empathy for pain on our skin. *Frontiers in Psychology*, *2*, 108.

Franklin, M., Ashton, H., Awad, E., & Lagnado, D. (2022). Causal framework of artificial autonomous agent responsibility. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 276-284).

Gombolay, M. C., Gutierrez, R. A., & Clarke, S. G.(2015). Decisionmaking authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots 39*:293–312

Gong, J. P., Zhang, X., Zhang, G., Wang, Q. L., & Fan, C. X. (2021). The relationship between college students' subjective well-being, coping styles, and social desirability. *Chinese Journal of Health Psychology, 29*(08), 1262-1265.

[巩金培, 张希, 张改, 王庆林 & 范成香. (2021). 大学生主观幸福感、应对方式与社会赞许性的关系. *中国健康心理学杂志 29* (08), 1262-1265.]

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior*, *119*.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd Ed.)*. New York: Guilford Press.

He, H., Martinsson, P., & Sutter, M. (2012). Group decision making under risk: An experiment with student couples. *Economics Letters, 117*(3), 691–693.

Hewstone, M., Rubin M., & Willis H. (2002). Intergroup bias. *Personality and Social Psychology Bulletin*, *53*, 575–604.

Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *Human–Computer Interaction, 19*(1–2), 151–181.

Hong, J. W., Cruz, I., & Williams, D. (2021). AI, you can drive my car: how we evaluate human drivers vs. self-driving cars. *Computers in Human Behavior, 125*(2), 106944.

Huang., L., Yang., Y. Z., & Ji., Z. M. (2003). Applicability of the Positive and Negative Affect Scale in Chinese. *Chinese Mental Health Journal, 17*(1), 54–56.

[黄丽, 杨廷忠, 季忠民. (2003). 正性负性情绪量表的中国人群适用性研究. *中国心理卫生杂志*, *17*(1), 54–56.]

Kim., T., & Hinds., P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In: Robot and human interactive communication, 2006. ROMAN 2006. The 15th IEEE international symposium on. IEEE, pp 80–85

Lei, X., & Rau, P. P. (2021). Effect of relative status on responsibility attributions in human–robot collaboration: Mediating role of sense of responsibility and moderating role of power distance orientation, *Computers in Human Behavior*, *122*,106820

Lei, X., & Rau, P.P. (2021). Should I Blame the human or the robot Attribution within a human–robot group. *International Journal of Social Robotics*, *13*(2), 363–377.

Levine M., Prosser A., Evans D., & Reicher S. (2005). Identity and emergency intervention: how social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*(4), 443–453.

Matsui, T., & Koike, A. (2021). Who Is to Blame? The Appearance of Virtual Agents and the Attribution of Perceived Responsibility. *Sensors, 21(8), 2646.*

Miller., D. T., & Ross., M. (1975). Self-serving biases in the attribution of causality: fact or fiction? *Psychological Bulletin*, *82*:213

Mittal, S. K. (2022). Behavior biases and investment decision: Theoretical and research framework. Qualitative Research in Financial Markets, *14*(2), 213–228.

Oliveira, R., Arriaga, P., Correia, F., & Paiva, A. (2020).Looking beyond collaboration: Socioemotional positive, negative and task-oriented behaviors in human–robot group interactions. *International Journal of Social Robotics*, *12*(2), 505–518.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1996, November). My partner is a real dog: Cooperation with social agents. Paper presented at the meeting of the CSCW 96 Conference on Computer-Supported Cooperative Work. New York: ACM.

Posard, M. N., & Rinderknecht, R. G. (2015). Do people like working with computers more than human beings? *Computers in Human Behavior, 51*, 232–238.

Savela, N., Kaakinen, M., Ellonen, N., & Oksanen, A. (2021).Sharing a work team with robots: The negative effect of robot co-workers on in-group identification with the work team, *Computers in Human Behavior*, *115*, 106585.

Silva, K., Chein, J., & Steinberg, L. (2020). The influence of romantic partners on male risk-taking. *Journal of Social and Personal Relationships*, *37*(5), 1405–1415.

Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science, 8*(4), 379-386.

Tang, W. C. (2017). *Men and Women Working Together can make of job easier: the study on risk decisions made by two-person group of different genders* (the Degree of Master). Beijing University of Posts and Telecommunications.

[唐伟超. (2017). *男女搭配干活不累：异性组合的两人共同风险决策研究*(硕士学位论文). 北京邮电大学.]

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987) Rediscovering the social group: A self-categorization theory. *British journal of social psychology*, *26*, 347–348.

Wang, X., & Krumhuber, E. G. (2018). Mind perception of robots varies with their economic versus social function. *Frontiers in Psychology, 9*, 1230.

Wang, Z., Kuang, Y., Tang, H., Gao, C., Chen, A., & Chan, K. (2018). Are Decisions Made by Group Representatives More Risk Averse? The Effect of Sense of Responsibility. *Journal of Behavioral Decision Making, 31*(3), 311–323.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal Personal Social Psychology, 54*(6), 1063–1070.

Wen, Z. L., & Ye, B. J. (2014). Different methods for testing moderated mediation models: Competitors or backups? *Acta Psychologica Sinica, 46*(5), 714–726.

[温忠麟, 叶宝娟. (2014). 有调节的中介模型检验方法: 竞争还是替补？*心理学报*, *46*(5), 714–726.]

Xiong, W., Wang, C., & Ma, L. (2023). Partner or subordinate? Sequential risky decision-making behaviors under human-machine collaboration contexts. *Computers in Human Behavior, 139*, 107556.

Xu, S. H., Fang, Z., & Rao, H. Y. (2013). Real or Hypothetical Monetary Rewards Modulates Risk Taking Behavior. *Acta Psychologica Sinica, 45*(8), 874–886.

[徐四华, 方卓, 饶恒毅. (2013). 真实和虚拟金钱奖赏影响风险决策行为. *心理学报*, *45*(8), 874–886.]

Xu, S. H., Pan, Y., Qu, Z., Fang, Z., Yang, Z. j., Yang, F., ... Rao, H. Y. (2018). Differential effects of real versus hypothetical monetary reward magnitude on risk-taking behavior and brain activity. *Scientific Reports, 8*(1), 3712.

Yue, L. Z., Li, S., & Liang, Z. Y. (2018). New avenues for the development of domain-specific nature of risky decision making. *Advances in Psychological Science, 26*(5), 928–938.

[岳灵紫, 李纾, 梁竹苑. (2018). 风险决策中的领域特异性. *心理科学进展*, *26*(5), 928–938.]

Yue, B., & Li, H. (2023). The impact of human-AI collaboration types on consumer evaluation and usage intention: a perspective of responsibility attribution. *Frontiers in Psychology, 14*,1277861.

You, S., Nie, J., Suh, K., Sundar, S.S. (2011). When the robot criticizes you…: self-serving bias in human-robot interaction. In: Proceedings of the 6th international conference on human–robot interaction. ACM, pp 295–296

Zhao, S. (2004). *A study on the relationship between Sensation seeking (SS) and mental health of college students*(the Degree of Master).Northeast Normal University, Changchun.

[赵闪. (2004). *大学生感觉寻求及其与心理健康关系的研究*(硕士学位论文). 东北师范大学, 长春.]

Human-AI cooperation makes individuals more risk seeking: The mediating role of perceived agentic responsibility

GENG Xiaowei[1,2], LIU Chao[3], SU Li[1], HAN Bingxue[1], ZHANG Qiaoming[4], WU Mingzheng[5]

([1] Department of Psychology, Hangzhou Normal University, Hangzhou 311121, China)

([2] Zhejiang Philosophy and Social Science Laboratory for Research in Early Development and Childcare, Hangzhou Normal University, Hangzhou 311121, China)

([3] School of Marxism, Binzhou Polytechnic, Binzhou 256600, China)

([4] College of Education, Ludong University, Yantai 264025, China)

([5] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou 311121, China)

Abstract

Risk decision-making involves choices made by individuals when they are uncertain about future outcomes. With advancements in artificial intelligence (AI), AI can now assist humans in making decisions. For instance, human drivers and AI drivers cooperate to carry out driving tasks, human doctors and AI doctors can collaborate on medical decisions. Currently, it is unclear how AI affects individuals' risk decision-making during such collaborations, which is crucial for enhancing the quality of human-AI decision-making. Therefore, studying the impact of human-AI cooperation on individuals' risk decision-making is essential.

In Experiment 1a, a total of 100 participants were recruited from one university. Employing a within-subject design, the independent variable was the partner type (i.e., human-human cooperation, human-AI cooperation, or no partner), while the dependent variable measured individuals' risk decision-making using the Balloon Analogue Risk Task (BART). In Experiment 1b, a total of 151 participants were recruited from another university and randomly assigned to two conditions: human-human cooperation and human-AI cooperation. As in Experiment 1a, the dependent variable remained the same. To investigate the mediating role of individual agentic responsibility, Experiment 2 recruited 199 participants from a university. This experiment utilized a between-subjects design, with the independent variable being the partner type (i.e., human-human cooperation or human-robot cooperation). Individual agentic responsibility was assessed by measuring the extent to which participants assumed responsibility for their tasks, and the dependent variable was individuals' risk levels as measured by the BART. Experiment 3 further explored the moderating effect of outcome

feedback. Participants received feedback based on their BART performance in Experiment 2, categorized as success or failure, and then assessed their perceived agentic responsibility before completing the BART again.

The results of Experiment 1a and 1b showed that participants in the control group (i.e., without cooperation) exhibited the highest risk-taking behavior, while those engaged in human-AI cooperation took greater risks than those in human-human cooperation. Results from Experiment 2 demonstrated that individual agentic responsibility partially mediated the effect of human-AI cooperation on individuals' risk decision-making. Specifically, participants reported a higher sense of agentic responsibility in human-AI cooperation compared to human-human cooperation, which contributed to increased risk-taking. Experiment 3 revealed that outcome feedback significantly moderates the mediating role of individual agentic responsibility regarding the influence of human-AI cooperation (versus human-human cooperation) on individuals' risk decision-making. Notably, under success conditions, participants attributed greater responsibility to themselves in human-AI collaboration compared to human-human collaboration. Conversely, under failure conditions, there was no significant difference in responsibility attribution between the two types of collaboration.

This research demonstrates that collaboration with AI can enhance an individual's propensity for risk-taking. Moreover, the influence of human-AI cooperation, compared to human-human cooperation, on individuals' risk decision-making is mediated by a sense of individual agentic responsibility and moderated by outcome feedback. These findings offer significant theoretical insights. Furthermore, this study holds substantial practical implications by aiding individuals in understanding how collaboration with AI impacts their risk-taking behaviors.

Keywords human-AI cooperation, human-human cooperation, risk decision making, perceived agentic responsibility, outcome feedback